Claims:

We claim.

1. A computer-based method for identifying invariant peptide motifs useful as drug targets wherein the said method comprises the steps of:

i) generating computationally overlapping peptide libraries from all the protein sequences of the selected organisms available at http://www.ncbi.nlm.nih.gov,

ii) sorting computationally the peptides of length 'N' obtained as above, alphabetically, according to single letter amino acid code,

iii) matching computationally common peptide sequences of the selected bacteria,

iv) locating computationally these common peptides in the original proteins and subsequently labeling them with their origin and location,

v) joining computationally the overlapping common peptides to obtain a long chain of invariant peptide sequences,

vi) annotating secondary structure of these conserved peptides from the crystal structure database,

vii) comparing pathogenic strain genomes against genomes of non-pathogenic strains and selecting the sequences not commonly conserved in these two groups,

viii) validating computationally the invariant sequence motifs as potential drug target sequence by searching for the given conserved sequences in the host genome and rejecting the ones present in the host genome.

2. The method of claim 1 wherein the length of the sliding window of length 'N' ranges from 4 to any length of amino acid residues.

3. The method of claim 1 wherein the protein sequence data is taken from any organism but not specifically limited to microbes such as *Mycoplasma pneumoniae*, *Helicobacter pylori*, *Hemophillus influenzae*, *Mycobacterium tuberculosis*, *Mycoplasma genitalium*, *Bacillus subtillis*, *Escherichia coli*.

4. A method as claimed in claim 1 where conserved peptide motifs as identified comprising:

1. AAQSIGEPGTQLT
2. AGDGTTTAT
3. AGRHGNKG
4. AHIDAGKTTT
5. CPIETPEG
6. DEPSIGLH
7. DEPTSALD
8. DEPTTALDVT
9. DHAGIATQ
10. DHPHGGGEG
11. DLGGGTFD
12. DVLDTWFSS
13. ERERGITI
14. ERGITITSAAT
15. ESRRIDNQLRGR
16. FSGGQRQR
17. GEPGVGKTA
18. GFDYLRDN
19. GHNLQEHS
20. GIDLGTTNS
21. GINLLREGLD
22. GIVGLPNVGKS
23. GKSSLLNA
24. GLTGRKIIVDTYG
25. GPPGTGKTLLA
26. GPPGVGKT
27. GSGKTTLL
28. GTRIFGPV
29. IDTPGHVDFT
30. IIAHIDHGKSTL
31. INGFGRIGR
32. IREGGRTVG
33. IVGESGSGKS
34. KFSTYATWWI
35. KMSKSKGN
36. KMSKSLGN
37. KNMITGAAQMDGAILVV
38. KPNSALRK
39. LFGGAGVGKTV
40. LGPSGCGK
41. LHAGGKFD
42. LIDEARTPLIISG
43. LLNRAPTLH
44. LPDKAIDLIDE
45. LPGKLADC
46. LSGGQQQR
47. MGHVDHGKT
48. NADFDGDQMAVH
49. NGAGKSTL
50. NLLGKRVD
51. NTDAEGRL
52. PSAVGYQPTLA
53. QRVAIARA
54. QRYKGLGEM
55. RDGLKPVHRR
56. SALDVSIQA
57. SGGLHGVG
58. SGSGKSSL
59. SGSGKSTL
60. SVFAGVGERTREGND
61. TGRTHQIRVH
62. TGVSGSGKS
63. TLSGGEAQRI
64. TNKYAEGYP
65. TPRSNPATY
66. VEGDSAGG
67. VRKRPGMYIG

5. A method as claimed in claim 1 wherein the number of invariant peptides varies according to the relatedness among the organisms and the number of organisms being compared.

6. A method as claimed in claim 1-4 wherein the invariant sequences belong to following proteins as available in the database http://www.ncbi.nlm.nih.gov wherein the said list of proteins comprise:

I DNA DIRECTED RNA POLYMERASE-BETA CHAIN

II EXCINUCLEASE ABC SUBUNIT A

III EXCINUCLEASE ABC SUBUNIT B

IV DNA GYRASE SUBUNIT B

V ATP SYNTHASE BETA CHAIN

VI S-ADENOSYLMETHIONINE SYNTHETASE

VII GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE

VIII ELONGATION FACTOR G (EF-G)

IX ELONGATION FACTOR TU (EF-TU)

X 30S RIBOSOMAL PROTEIN S12

XI 50S RIBOSOMAL PROTEIN L12

XII 50S RIBOSOMAL PROTEIN L14

XIII VALYL tRNA SYNTHETASE (VALRS)

XIV CELL DIVISON PROTEIN FtSH HOMOLOG

XV DnaK PROTEIN (HSP70)

XVI GTP BINDING PROTEIN LepA

XVII TRANSPORTER

XVIII OLIGOPEPTIDE TRANSPORT ATP BINDING PROTEIN OPPF

7. A method as claimed in claim 1 wherein the said method of comparing the peptide libraries as given in step (iii) of claim 1 is carried out by following the steps given in figure 1.
8. A method as claimed in claim 1 wherein the said method of locating the common peptides in the original protein sequences as given in step (iv) of claim 1 is carried out by following the steps given in figure 2.
9. A method as claimed in claim 1 wherein the said method of creating a common peptide of variable length after removing the overlappings as given in step (v) of claim 1 is carried out by following the steps given in figure 3.
10. A microprocessor based system for performing the methods of the invention which comprises:

i) means of determining the amino acid sequence window for creation of peptide library and subsequent origin tagging,

ii) means of comparing the peptide library,

iii) locating computationally these common peptides in the original proteins and subsequently labeling them with their origin and location,

iv) joining computationally the overlapping common peptides to obtain a long chain of invariant peptide sequences,

11. A computer based system for performing the methods of the invention further comprising a central processing unit, executing peptide library creating program (PEPLIB), peptide library matching program (PEPLIMP), peptide stitching program (PEPSTITCH), peptide extraction program (PEPXTRACT) wherein the said programs are all stored in a memory device accessed by the central processing unit connected to a display on which the central processing unit displays the screens of the above mentioned programs in response to user inputs with a user interface device.

12. A method for assigning function to a protein of unknown function showing no/weak homology to other protein sequences in a publicly available database (SWISSPROT) by employing the following steps:

I. generating computationally overlapping peptide library from the protein sequences of unknown function,

II. sorting computationally the peptides of length 'N' (N is the length of the sliding window of amino acids) obtained as above, alphabetically, according to single letter amino acid code,

III. matching computationally the current library with peptide library of all functionally known proteins to obtain common peptides,

IV. locating computationally these common peptides in the original proteins and subsequently labeling them with their origin and location,

V. joining computationally the overlapping common peptides to obtain a long chain of invariant peptide sequences,

VI. assigning function to the unknown protein based on the function of the protein with which maximum length of peptide sequence identity is found. The more is the number of matches with the proteins of similar function the likelihood of functional assignment will be higher.